BioPartsDB

Bioinformatics, 32(22), 2016, 3519–3521
doi: 10.1093/bioinformatics/btw394
Advance Access Publication Date: 13 July 2016
Applications Note

Systems biology

BioPartsDB: a synthetic biology workflow web-application for education and research

Giovanni Stracquadanio[1], Kun Yang[1], Jef D. Boeke[2] and Joel S. Bader[1,*]

[1]Department of Biomedical Engineering and High-Throughput Biology Center, Johns Hopkins University, Baltimore, MD 21218, USA and [2]Institute for Systems Genetics and Department of Biochemistry and Molecular Pharmacology, NYU Langone Medical Center, New York, NY 10016, USA

*To whom correspondence should be addressed.

Associate Editor: Janet Kelso

Received on September 24, 2015; revised on February 16, 2016; accepted on June 17, 2016

Abstract

Summary: Synthetic biology has become a widely used technology, and expanding applications in research, education and industry require progress tracking for team-based DNA synthesis projects. Although some vendors are beginning to supply multi-kilobase sequence-verified constructs, synthesis workflows starting with short oligos remain important for cost savings and pedagogical benefit. We developed BioPartsDB as an open source, extendable workflow management system for synthetic biology projects with entry points for oligos and larger DNA constructs and ending with sequence-verified clones.

Availability and Implementation: BioPartsDB is released under the MIT license and available for download at https://github.com/baderzone/biopartsdb. Additional documentation and video tutorials are available at https://github.com/baderzone/biopartsdb/wiki. An Amazon Web Services image is available from the AWS Market Place (ami-a01d07c8).

Contact: joel.bader@jhu.edu

1 Introduction

Synthetic biology projects require software systems to organize, track and verify the creation of parts encoded by DNA sequences, with individual parts corresponding to genomic elements such as promoters or coding domains. Part sizes can range from 100 nucleotide regulatory elements to megabase chromosomes. Tasks include organizing an experimental plan, tracking protocols, retaining quality control assessments, assessing performance and storing a verified part in a repository.

Biological enthusiasts, such as DIY Bio (http://diybio.org), educational initiatives such as iGEM (https://www.igem.org), and new synthetic biology courses such as Build-a-Genome (Cooper *et al.*, 2012) all require workflow systems to match to DNA synthesis protocols. Protocols depend on whether a DNA construct or part is obtained by PCR amplification, restriction digest or direct synthesis, and direct synthesis can involve different hierarchical assembly strategies. Quality control procedures and parts storage, whether in a vector or as DNA, can also depend on the synthetic route. While laboratory information management systems (LIMS) are available for synthetic biology (Ham *et al.*, 2012), they focus primarily on synthetic DNA inventory management rather than on the synthesis process itself. We therefore designed BioPartsDB as an open source, modular tool for DNA synthesis workflow and parts registration, providing a ready resource for research and education. Our permissive license makes the software fully available for industrial and commercial applications as well.

2 Results and discussion

2.1 Overall process and unit operations

Parts are constructed through a series of unit operations (Fig. 1). Each operation has well-defined characteristics: (i) input reagents; (ii) output products; (iii) status (e.g. pending, done); (iv) quality control metrics; and (v) a protocol for converting the input to output and performing quality assessment (QA) and quality control (QC). These must be fully specified at each step in the process, and a successful QC, termed a 'Pass', is required for the products to advance to the next step.

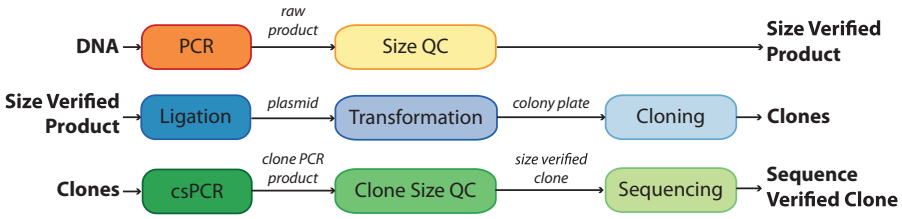

Fig. 1. BioPartsDB workflow

2.2 Target design and synthetic plan

Two user roles are implemented: *admin* and *members*. While members can only run experiments, admin users are responsible for loading parts data into the system, managing sequencing plates, updating protocols and managing users. Admins assign parts to the members for synthesis. Target sequences can be designed using design tools, such as BioPartsBuilder (http://public.biopartsbuilder.org), which is compatible with BioPartsDB.

2.3 PCR modules: synthesis plan to raw product

BioPartsDB enables two distinct routes from a synthesis plan to a raw product: standard amplification from genomic DNA (sPCR), and templateless PCR (tPCR) from overlapping oligos followed by finish PCR (fPCR) of a target amplicon. 'Templeteless PCR' and 'Finish PCR' are nomenclature for the two steps in a polymerase chain reaction (PCR) adopted from the Build-a-Genome course, in which students learn synthetic biology by taking part in entire chromosome synthesis projects (Cooper *et al.*, 2012). The fPCR output is a raw PCR product. The sPCR, tPCR and fPCR modules share a common interface for usability. Each module has a home page showing the workflow status (pending, finished) and available QC reports for each experiment to assist with troubleshooting problems including faulty PCR machines or reagents. The modules assist reaction set-up by providing a map of source wells for input oligo plates and computing reagent volumes for master mixes.

2.4 Size QC module: raw product to size-verified product

Sizing is performed by gel electrophoresis, the first QC step for the part itself. Parts must pass to proceed. A user's module home page lists that user's gels, each assigned a sequential number by the software. For each gel, the system displays the gel image, the loading order and the QC value for each lane with a raw product. The user can prepare a new gel by selecting PCR products of any kind; the system will show a report page with the list of PCRs to load on the gel and the loading order, which can be changed through a drag-and-drop interface. Selecting the gel image permits the user to label each lane with the corresponding PCR product. The system generates an image map and, when moving over a lane, displays the name of the PCR product and the expected length. This feature is useful for presentations and in-class discussions.

2.5 Ligation module: sized product to plasmid

Size-verified PCR products are introduced into vectors to create plasmids. Ligation, Gibson assembly (Gibson *et al.*, 2009), terminal homologous recombination and TA cloning are examples of alternatives that are compatible with BioPartsDB. When creating a new ligation experiment, the user selects a PCR product and the system displays the protocol and the master mix aliquots for each reagent. The ligation must be marked as 'Pass' to proceed to transformation.

2.6 Transformation module: plasmid to colony plate

Transformation introduces ligation products into a bacterial, yeast, mammalian or other host cell. A new transformation is initiated by specifying a ligation product and host strain; the system generates experimental identifiers that can be reported on the plate for tracking. The transformation module supports blue–white colony screens used for bacterial transformations to identify transformants and assess transformation efficiency. The primary data collected are counts of blue and white colonies, and a QC value is reported based on successful recovery of transformant colonies. The blue–white screening step is easily removed if desired.

2.7 Cloning module: colony plates to clones

The cloning module assigns unique identifiers to each colony and generates maps for growth plates. The input reagents for a new batch of clones are the source transformations from which colonies will be picked. The software reports a summary of the operation. Each clone is given a unique identifier from an auto-incrementing index for each part. The system also generates a 96-well plate map specifying the locations of colonies during growth; hardcopies of plate maps may be placed underneath plates for easy colony picking. As a final step, the colonies are marked as 'Picked' to proceed to csPCR.

2.8 csPCR module: clone to PCR product or restriction digest

The colony screening PCR module takes clones as input, rather than PCR products or other nucleic acids, and is designed for 96-well PCR plates. Clones may be selected from growth plates corresponding to multiple parts. The system then generates a PCR plate map showing the protocol along with the aliquots for a master mix. After running a csPCR, the user reports the experiments as 'Finished', or 'Fail' for an obvious error. The current software supports PCR but could be modified to support sizing by restriction digest fragment pattern instead.

2.9 Clone size QC: PCR or restriction product to size-verified clone

The csPCR gel module requires csPCR products as input and is otherwise similar to the PCR gel module. csPCR gels are the final pre-sequencing QC; parts marked as 'Pass' may be sent for sequencing.

2.10 Sequencing module: size-verified clone to sequence-verified clone

Inputs for the sequencing module are plates of clones, which may correspond to multiple lab members and parts. The module tracks the number of clones than can be sent, the number of wells available and the clone locations. Clones that have passed csPCR QC are picked from growth plates. The system then generates a plate map for re-arraying the clones from growth plates onto a single sequencing plate. With this design, lab members can independently prepare

clones for sequencing and still take advantage of large sequencing batches. Quality control values, 'Pass' for error-free clones, from expert or automated validation analysis (Lee *et al.*, 2010) can be reported through CSV files by an administrator.

2.11 Administration

The BioPartsDB system provides administrative module that are responsible for managing parts, oligo plates, vendors, lab materials (strain, vectors), devices (PCR machines, gel boxes), sequencing plates, protocols, users and groups.

Funding

This work was supported by National Science Foundation [grant number MCB-0718846] and National Institutes of Health [grant number P50GM107632] to J.D.B. and National Science Foundation [grant number MCB-1445545] to J.S.B.

Conflicts of Interest: none declared.

References

Cooper,E.M. *et al.* (2012) The build-a-genome course. *Methods Mol. Biol.*, **852**, 273–283.

Gibson,D.G. *et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343–345.

Ham,T.S. *et al.* (2012) Design, implementation and practice of jbei-ice: an open source biological part registry platform and tools. *Nucleic Acids Res.*, **40**, e141–e141.

Lee,P.A. *et al.* (2010) Cloneqc: lightweight sequence verification for synthetic biology. *Nucleic Acids Res.*, **38**, 2617–2623.